MONOCULAR OBJECT LOCALIZATION BY SUPERQUADRICS CURVATURE REPROJECTION AND MATCHING

Enrico Zappia, Ilya Afanasyev, Nicolo’ Biasi, Mattia Tavernini, Alberto Fornaser, Antonio Selmo and Mariolino De Cecco
Abstract

We present a new method for 3D object localization from a single image. The main new idea is to match 2D image gradient to the reprojection of 3D curvature to retrieve objects position relative to the camera. The object parameters are a-priori known and modelled by SuperQuadrics (SQ) that enable the calculation of the analytical form of curvature. The image processing stage includes object detection and segmentation by the Histogram of Oriented Gradients (HOG) algorithm. The method proposed uses the dependencies between SQ curvature and image gradient also considering the illumination model and object contour embedded in a proper cost function. To manage local minima we propose the use of particle swarm optimization (PSO).
Problem of localization

Detection

Pose estimation
The main idea is to reproject a model contour on the gradient image. The method is very time efficient, but the object’s profile is often strongly varied along the edges due to e.g. clutter, shading, and texture.
Region based


Region based rely on the homogeneity of spatially localized features. This assumption may not always be verified.
Related Works

Range Data

G. Biegelbauer and M. Vincze. Efficient 3d object detection by fitting superquadrics to range image data for robot’s object manipulation.

Ilya Afanasyev, Massimo Lunardelli et. al; 3d human body pose estimation by superquadrics.

Range data techniques do not suffer from many image problems, but deal with wider dataset and requires particular acquisition technology based on 3D sensors or multicamera setup.
In this work is proposed a novel method to fuse region based and edge matching techniques providing a very compact formulation, using curvature reprojection.

<table>
<thead>
<tr>
<th></th>
<th>Clutter</th>
<th>Occlusion</th>
<th>Texture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edge Matching</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Region Based</td>
<td>P</td>
<td>N</td>
<td>VP</td>
</tr>
<tr>
<td>Range Data</td>
<td>P</td>
<td>VP</td>
<td>-</td>
</tr>
<tr>
<td>Curvature reproj.</td>
<td>P</td>
<td>VP</td>
<td>N</td>
</tr>
</tbody>
</table>

N-negative  P-positive  VP-very positive

Curvature is used in closed form due to parametric formulation of objects models.
Superquadrics are mathematical functions allowing the representation of a pretty high number of elementary solids.

\[ F(x, y, z) = \left( \left( \frac{x}{a_1} \right)^{2/\varepsilon_2} + \left( \frac{y}{a_2} \right)^{2/\varepsilon_2} \right)^{\frac{\varepsilon_2}{\varepsilon_1}} + \left( \frac{z}{a_3} \right)^{2/\varepsilon_1} \]

where \( x, y, z \) - SQ coordinate system
\( a_1, a_2, a_3 \) - scale parameters of the object;
\( \varepsilon_1, \varepsilon_2 \) - object shape parameters;
SuperQuadrics Models

Due to implicit formulation it is possible to evaluate close form mean curvature and normal directions:

- **Normal directions:**

\[
\mathbf{N}(x, y, z) = \frac{\nabla \mathbf{F}(x, y, z)}{|\nabla \mathbf{F}(x, y, z)|}
\]

- **Mean curvature:**

\[
K_M = \frac{\nabla \mathbf{F} \cdot H(\mathbf{F}) \cdot \nabla \mathbf{F}^T - |\nabla \mathbf{F}|^2 \text{Trace}(H)}{2 |\nabla \mathbf{F}|^3}
\]
Superquadrics models:

Curvature is mapped in grayscale colors.

Enrico Zappia, Ilya Afanasyev, et.al. – GraphiCon’2012 – Moscow, Russia – 03/10/2012
Objects Detection is performed with HOG (Histogram of Oriented Gradients).

From detection data PSO (particle swarm optimization) optimizer is initialized.

Cost function is a combination of:
- Curvature/gradient matching
- Light model convolution
- Normals/gradient alignment

The optimization is continued until the convergence.
HOG is an algorithm that uses the distribution of intensity gradients for object detection.

HOG’s finale results is a bounding box around the detected object.
From HOG detection, inverting the calibration formula, is possible to switch from pixel coordinate \((X_n)\) to world coordinate \((X_w)\). Anyway due to perspective annihilation depth data is lost.

According to detection window area \((A)\) an indicative guess can be done.

\[
X_w = \frac{k}{A} \begin{bmatrix} x_n \\ y_n \\ 1 \end{bmatrix}
\]

So trust region is sets:

<table>
<thead>
<tr>
<th>(\theta_z)</th>
<th>(X_w)</th>
<th>(Y_w)</th>
<th>(Z_w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\pi/2)</td>
<td>(X_w -</td>
<td>X_w</td>
<td>\cdot 0.3)</td>
</tr>
<tr>
<td>(\pi/2)</td>
<td>(X_w +</td>
<td>X_w</td>
<td>\cdot 0.3)</td>
</tr>
</tbody>
</table>
Cost Function-Curvature/Gradient

The idea is to evaluate the euclidean distance \( E_f \) from normalized gradient \( G^N \) and mean curvature \( K^N \).

\[
E_f = \frac{\sum_{i=1}^{n} ||G^N(i) - K^N(i)||^2}{n} + ||G^N(i) - K^N(i)||^\infty
\]
Using Phong’s model, the function score is obtained with the convolution between the digital lighting image ($I_p^N$) and the real gray scale image ($L_g^N$).

\[ C_f = \sum_{i=1}^{n} \frac{I_p^N(i) \cdot L_g^N(i)}{||I_p^N||^2 \cdot ||L_g^N||^2} \]
Cost Function-Normals/Gradient Align

If the matching is good the reprojected normals on the image plane are almost parallel to silhouette’s gradients. Quality of matching is estimated counting the number of approximately parallel vectors:

\[ V_f = \frac{1}{n} \sum_{i=1}^{n} \text{If}(\left| \cos(\theta_{NG}(i)) \right| > 0.9, 1) \]

Silhouette extraction      Normals eval.      Normals reprojection
The final cost function is the ratio of three contributions.

\[
F_c = \frac{\sum_{i=1}^{n} \|G^N(i) - K^N(i)\|^2}{\sum_{i=1}^{n} \frac{I_p^N(i) L_g^N(i)}{\|I_p^N\|^2 \cdot \|L_g^N\|^2}} + \frac{\|G^N(i) - K^N(i)\|_\infty}{\frac{1}{n} \sum_{i=1}^{n} \mathbf{1f}(\cos(\theta_{NG}(i)) > 0.9, 1)}
\]
The problem is multi-modal, so common gradient based techniques are not suitable.

Using a multi-agent cooperative genetic algorithm, the solutions will be:

\[
\mathbf{v}_{k+1}^{(s)} = \omega \cdot \mathbf{v}_{k}^{(s)} + C_1 r_1 (\mathbf{p}_{k}^{(s)} - \mathbf{x}_{k}^{(s)}) + C_2 r_2 (\mathbf{g}_{k} - \mathbf{x}_{k}^{(s)}) \]

\[
\mathbf{F}_{k}^{(s)}(\mathbf{p}_{k}^{(s)}) \quad \text{Cognitive Term} \quad \mathbf{F}_{k}^{(s)}(\mathbf{g}_{k}) \quad \text{Social Term}
\]
Calibration of the method

We aim to prove detection correctness and a quantitative estimation of the localization accuracy with state-of-art algorithms that use RGB-D camera.
MS-Kinect Localization

Kinect localization is performed with an accurate filtering of 3D data. The result is a set of clustered points cloud with no outliers. Then fitting is performed with Levenberg-Marquardt regression algorithm.
The ellipses are estimated to quantify uncertainty using eigenvalues along principal directions for each ellipse (green - for Kinect, blue - for our method).

<table>
<thead>
<tr>
<th></th>
<th>Cuvature [mm]</th>
<th>MS-Kinect [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOX</td>
<td>$\lambda_1 = 18.85$ $\lambda_2 = 27.23$</td>
<td>$\lambda_1 = 11.65$ $\lambda_2 = 40.56$</td>
</tr>
<tr>
<td>CYLINDER</td>
<td>$\lambda_1 = 11.35$ $\lambda_2 = 13.55$</td>
<td>$\lambda_1 = 13.83$ $\lambda_2 = 47.05$</td>
</tr>
</tbody>
</table>
Results-General pose

These figures present some localization results for objects in general positions and in occluded situations.
Results-Without light informations

Here are proposed some without light informations, the sistem works but is more affected by minimal local problems.
Conclusion

The article presents an innovative approach for monocular object localization. This approach fuses (1) curvature reprojection to gradient image; (2) light source information; (3) silhouette’s normals orientation.

Our method gives an object localization accuracy comparable and in some cases even better than kinect localization.

<table>
<thead>
<tr>
<th>Cuvature [mm]</th>
<th>MS-Kinect [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOX</td>
<td>$\lambda_1 = 18.85 ; \lambda_2 = 27.23$</td>
</tr>
<tr>
<td>CYLINDER</td>
<td>$\lambda_1 = 11.35 ; \lambda_2 = 13.55$</td>
</tr>
</tbody>
</table>

The algorithm is competitive in cluttered scenes as it relies on the whole object in contrast to only edges as in the case of edge/gradient methods, and it provides a more compact formulation than region based approaches.
Future works and Improvement

- Optimization is still an open problem. The correct orientation retrieval may generate many minimal local problems. It is important to point that the lowest values of the cost function always lead to correct localization.

- Future works may deal with using more complicated models (composed SQ models).

- Using tracking to refine the pose estimation.

- Applying embed shape from shading techniques to retrieve curvature information about objects from the static image.
The authors are very grateful to colleagues from Mechatronics dep. (University of Trento) and EU grants, incl. Marie Curie-COFUND-Trentino postdoc program, 2010-2013.
Thanks for the attention
HOG is an algorithm that uses the distribution of intensity gradients for object detection.

HOG involves two main phases: learning and detection. During learning features are extracted and store into HOG descriptor.
Stored data are then processed with a Support Vector Machine (SVM).

The results is a Root Model used for further detection.
Results

The occurrence of errors for cylinders is less than 30 mm (1% maximum distance) for boxes little higher (2% maximum distance).
From the proposed algorithm's estimation, objects positions on grid are reconstructed: